Popovich et al. Supplementary Materials

Supplemental Methods

Multilocus sequencing typing results were assigned using ARIBA[1] (Supplemental Table 5). The software was upgraded using the annotation tool snpEff in SPANDx to version 4.1. Three SPANDx analyses were performed. Initially, all genomes were processed simultaneously using a USA300 reference genome (GenBank accession NC_010079). Subsequently, putative USA300 strains only (<1100 SNVs relative to NC_010079) were processed using the same USA300 reference genome. In addition, putative USA100 strains only (<700 SNVs relative to CP029474 and CP029475) were processed using a closed genome from this study (30 Pt nares) as described previously[2]. In all cases, the reference genome was normalized with Picard Normalize-Fasta module. The SPANDx data output consisted of a single-nucleotide variant (SNV) matrix of 42,597 positions for the first analysis of all genomes, 3,057 positions for analysis of USA300 strains only, and 3,598 positions for analysis of USA100 strains only. Mapping statistics are shown in Supplemental Tables 3 and 4 for USA100 and USA300 strains, respectively. Nucleotide differences were used as a measure of genetic distance between strains.

For comparative genome evaluations, an SNV matrix from all genomes was imported into the software package MEGA7 for phylogenetic analyses, including boot-strapped maximum likelihood and neighbor-joining trees. The SNV matrix was also used for Bayesian phylogenetic analysis using the software package

MrBayes[3]. Clusters of MRSA strains were identified when all 3 phylogenetic analyses indicated support for nodes, including bootstrap values of >70% for maximum likelihood and neighbor-joining analysis, and posterior probability values of >95% for Bayesian analysis. Subsequently, SNV matrices from USA100 and USA300 genome analyses alone were processed in MEGA7, generating boot-strapped neighbor-joining trees. Distance sales are in the units of numbers of base differences per sequence.

Supplemental references

1. https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000131

2. Chlipala GE, Lei ZA, Maienschein-Cline M, et al. Complete Genome Sequence of a USA100 Methicillin-Resistant Staphylococcus aureus Strain. Microbiology resource announcements 2019;8.

3. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572-4.

Supplemental Table

Table S1 – Summary of sampling sites by episode number and USA100, USA300, or non-USA100/300 designation

Table S2- Identified clusters on phylogenetic trees with associated single nucleotide variant distances when compared to USA300 or USA100 reference genomes.

Table S3- Percentage of mapped reads and mapping coverage for USA100 genomes as processed through the SPANDx pipeline using the USA100 reference (CP029474 and CP029475)

Table S4- Percentage of mapped reads and mapping coverage for USA300 genomes as processed through the SPANDx pipeline using the USA300 reference (NC_010079; TCH1516)

Table S5- Multilocus sequencing typing results for collected isolates

Supplemental Figures

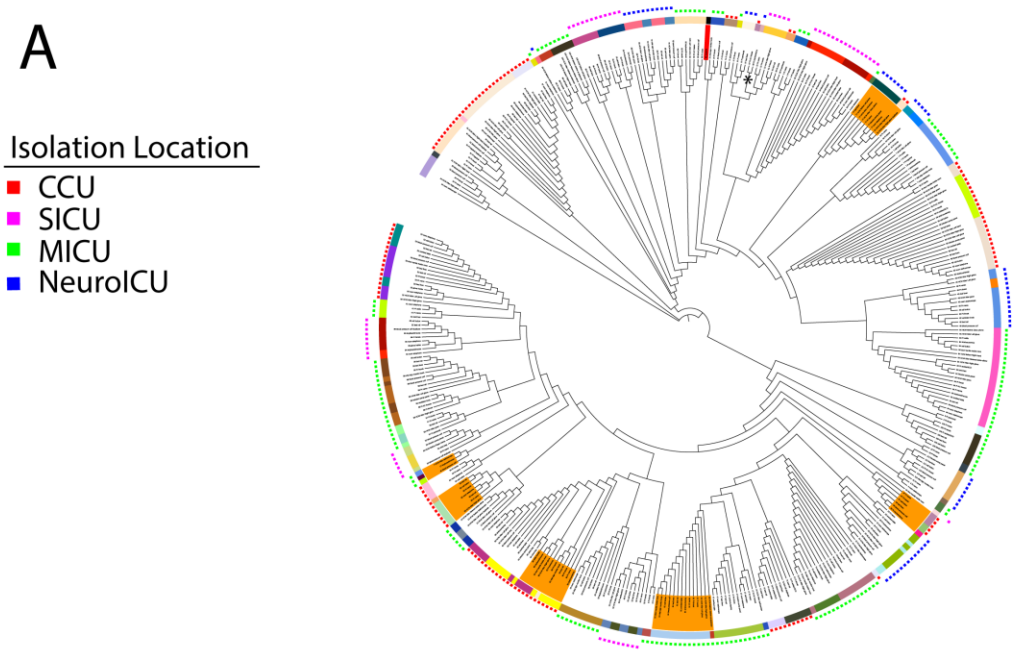
Supplemental Figure 1. Phylogenetic analysis for all 413 sequenced MRSA strains, relative to a USA300 reference genome. (A) A circular tree is used to demonstrate relationships between MRSA strains. Metadata for each isolate include an isolation source (identified in name), an episode number (identified by name and in color in the innermost ring) and hospital location (identified by color in the outermost ring). Statistically supported clusters (as defined in the manuscript) composed of isolates from encounters from different healthcare workers, different patients, or different environments are highlighted in orange. The tree is based on WGS of MRSA isolates, processed through a SPANDx pipeline, and distance matrix is the number of nucleotide variants between genomes. A total of 42,597 nucleotide positions were analyzed. The USA300 reference genome (NC_010079) is highlighted in red. One-hundred fifty-nine isolates were identified as USA300 strains, 214 were identified as USA100 strains, and the remaining 40 isolates were non-USA300, non-USA100 strains. **(B)** An unrooted maximum likelihood phylogenetic tree of all MRSA isolates based on WGS data. The scale bar represents 1000 nucleotide variants.

Supplemental Figure 2. Histogram of pair-wise comparison of SNV distance between all USA300 genome strains generated in this study. The x-axis represents number of SNVs between each pair of genomes, and the y-axis is the number of comparisons. The full distribution of comparisons is shown (top) as well as comparisons of genomes within 100 SNVs (bottom).

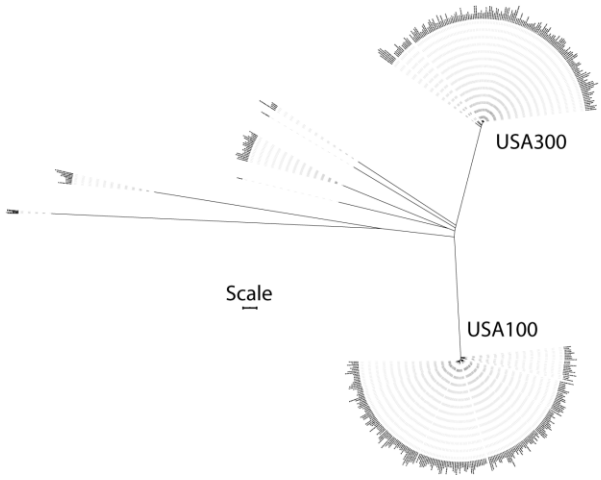
Supplemental Figure 3. Histogram of pair-wise comparison of SNV distance between all USA100 genome strains generated in this study. The x-axis represents number of SNVs between each pair of genomes, and the y-axis is the number of comparisons. The full distribution of comparisons is shown (top) as well as comparisons of genomes within 100 SNVs (bottom).

Supplemental Figure 1

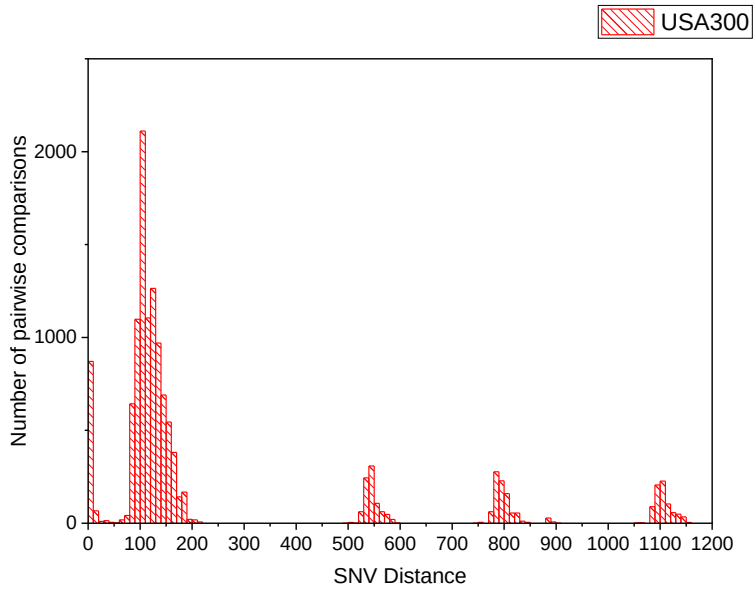
A

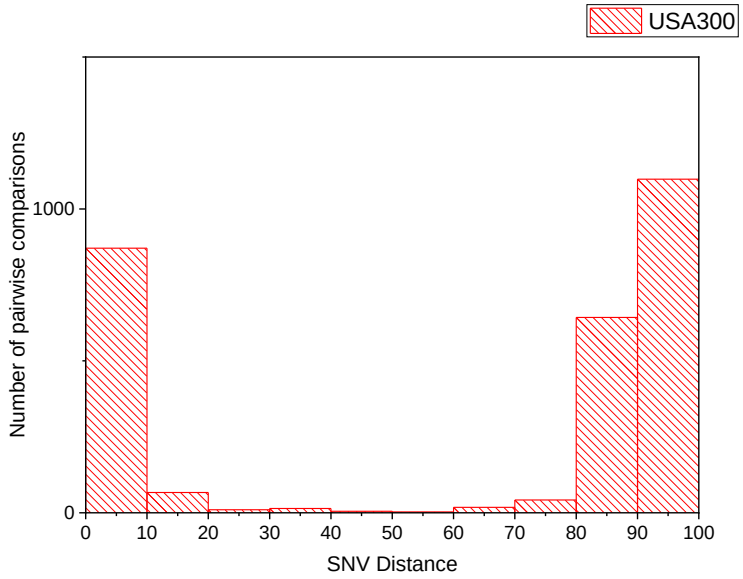


B



Supplemental Figure 2





Supplemental Figure 3